

Audio Information Transforming Method, Audio Information  
Transforming Program, and Audio Information Transforming  
Device

5

Background of the Invention

Field of the Invention

The present invention relates to an audio  
information transforming device, an audio information  
transforming method, and an audio information transforming  
10 program, which are employed in a video/audio format like  
MPEG (Moving Picture Coding Experts Group) 4 having video  
information and audio information every object, or a  
video/audio format like DVD (Digital Versatile Disk) having  
video information and audio information every scene.

15

Description of the Related Art

In recent years, the video streaming based on the  
DVD or the broadband is being prosperously carried out, and  
thus a chance to handle the video/audio format in the home  
is increased. In particular, since the DVD is spread and  
20 the audio apparatuses such as the AV amplifier, etc. become  
inexpensive, the persons who enjoy the audio in the  
multiple channels are increased. In the DVD, MPEG 2 is  
used as the video recording system and Dolby digital (AC-  
25 3), DTS (Digital Theater System), linear PCM (Pulse Code

Modulation), MPEG audio, or the like is used as the audio recording system. Eight audio streams can be installed into the DVD disk. Thus, if a different sound is loaded on each audio stream respectively, various applications such as dubbing of plural languages, high sound quality playing, commentary, sound track, etc. can be implemented.

Meanwhile, there is MPEG 4 as one of the next generation video/audio formats. In the MPEG 4, the object having video/audio information constituting the scenes that are reproduced on the screen is observed with interest, and the motion picture compression can be effectively attained by coding the motion picture every object.

However, in the multi-channel (e.g., 5.1-channel, etc.) audio system for playing the DVD in the prior art, it is impossible to change the listening point obtained by one audio stream. Therefore, the listener can get the hearing feeling only at the listening point at which the listener himself or herself listens to the audio.

Also, the audio information of each object are converted into the multi-channel audio, a quantity of data is increased. Therefore, a video/audio recording time is shortened correspondingly.

### **Summary of the Invention**

The present invention has been made in view of the

above circumstance, and it is an object of the present invention to provide an audio information transforming device, an audio information transforming method, and an audio information transforming program, which are capable  
5 changing a listening point freely only by one audio stream to thereby produce the audio environment that enables the listener to feel that such listener is just in the video, and also achieving multiple channels only by 1-channel audio information to thereby save a quantity of data and  
10 prolong a video/audio recording time.

In order to achieve the above object, an audio information transforming method set forth in Claim 1 provides an audio information transforming method applied to a video/audio format in which a screen contains a  
15 plurality of objects and each object has video information, position information, and audio information, which comprises a virtual listening point setting step of setting a virtual listening point at a position different from a basic listening point that is set as a position at which a  
20 listener listens to an audio; a comparing step of comparing a positional relationship between the basic listening point and the object with a positional relationship between the virtual listening point and the object; and a changing step of changing an allocation ratio of an audio to a plurality  
25 of audio outputting means based on a compared result in the comparing step.

According to such method, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced.

Also, an audio information transforming method set forth in Claim 2 provides an audio information transforming method applied to a video/audio format in which each scene produced on a screen has video information, audio information, and a virtual sound source, which comprises a setting step of setting a virtual listening point at a position different from a basic listening point that is set as a position at which a listener listens to an audio; a comparing step of comparing a positional relationship between the basic listening point and the virtual sound source with a positional relationship between the virtual listening point and the virtual sound source; and a

changing step of changing an allocation ratio of an audio to a plurality of audio outputting means based on a compared result in the comparing step.

According to such method, the positional relationship between the basic listening point and the virtual sound source is calculated and also the positional relationship between the virtual listening point and the virtual sound source is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced.

Also, an audio information transforming method set forth in Claim 3 provides an audio information transforming method applied to a video/audio format in which a screen contains a plurality of objects and each object has video information, position information, and 1-channel audio information, which comprises a step of calculating a positional relationship between a basic listening point, which is set as a position at which a listener listens to an audio, and the object; and a step of allocating the 1-channel audio information to a plurality of audio outputting means based on the positional relationship.

According to such method, the 1-channel audio information that the object has can be allocated to the multiple channels of the audio based on the positional relationship between the basic listening point and the object, and can be output to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system). Therefore, a quantity of data can be saved and the video/audio recording time can be prolonged.

Also, an audio information transforming method set forth in Claim 4 provides an audio information transforming method applied to a video/audio format in which a screen contains a plurality of objects and each object has video information, position information, and 1-channel audio information, which comprises a setting step of setting a virtual listening point at a position different from a basic listening point that is set as a position at which a listener listens to an audio; a comparing step of comparing a positional relationship between the basic listening point and the object with a positional relationship between the virtual listening point and the object; and an allocating step of allocating the 1-channel audio information to a plurality of audio outputting means based on a result in the comparing step.

According to such method, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to

the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the 1-channel audio information can be allocated to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (virtual listening point), can be produced, and also a quantity of data can be saved and the video/audio recording time can be prolonged.

In addition, in the audio information transforming method according to any one of claim 1, claim 2, or claim 4 as the audio information transforming method set forth in claims 5 to 7, the virtual listening point or the virtual sound source has direction information.

According to such method, the direction information is provided to the virtual listening point or the virtual sound source. Therefore, the listener can feel that such listener just enters into the video (virtual listening point), and the direction to which the listener looks can be changed, and the audio environment with the stronger appeal/reality can be produced.

In order to achieve the above object, an audio

information transforming program set forth in claim 8 for causing a computer to execute a procedure of setting a virtual listening point; a procedure of comparing a positional relationship between a basic listening point and an object with a positional relationship between the virtual listening point and the object; and a procedure of changing an allocation ratio of an audio to a plurality of audio outputting means based on a result in the comparing procedure.

10           In the reproduced audio having the video/audio format such as MPEG 4, for example, such program calculates the positional relationship between the basic listening point and the virtual sound source and also calculates the positional relationship between the virtual listening point and the virtual sound source, and then changes the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio environment with the strong appeal/reality, which permits the listener to feel that such listener just enters into the video (the virtual listening point), can be



implemented.

Also, an audio information transforming program set forth in claim 9 for causing a computer to execute a procedure of setting a virtual listening point; a procedure  
5 of comparing a positional relationship between a basic listening point and a virtual sound source with a positional relationship between the virtual listening point and the virtual sound source; and a procedure of changing an allocation ratio of an audio to a plurality of audio.  
10 outputting means based on a result in the comparing procedure.

In the video/audio format having the audio information and the video information in unit of the scene that is reproduced on the screen, such program calculates  
15 the positional relationship between the basic listening point and the virtual sound source and also calculates the positional relationship between the virtual listening point and the virtual sound source, and then changes the allocation ratio of the audio to a plurality of audio  
20 outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player  
25 (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio

environment with the strong appeal/reality, which permits the listener to feel that such listener just enters into the video (the virtual listening point), can be implemented.

5           Also, an audio information transforming program set forth in claim 10 for causing a computer to execute a procedure of calculating a positional relationship between a basic listening point and an object; and a procedure of allocating 1-channel audio information to a plurality of  
10 audio outputting means based on the positional relationship.

          In the reproduced audio having the video/audio format such as MPEG 4, for example, such program allocates the 1-channel audio information that the object has to the  
15 multiple channels of the audio based on the positional relationship between the basic listening point and the object, and then outputs such audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system). Therefore, if the recording medium (the memory  
20 such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can save a quantity of data and prolong the video/audio recording time, can be implemented.

25           Also, an audio information transforming program set forth in claim 11 for causing a computer to execute a

procedure of setting a virtual listening point; a procedure of comparing a positional relationship between a basic listening point and an object with a positional relationship between the virtual listening point and the object; and a procedure of allocating 1-channel audio information to a plurality of audio outputting means based on a result in the comparing procedure.

In the reproduced audio having the video/audio format such as MPEG 4, for example, such program decides the virtual listening point at the position that is different from the basic listening point, then calculates the positional relationship to the virtual listening point based on the position information that the object has and also calculates the positional relationship between the object and the basic listening point, and then allocates the 1-channel audio information to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), and also can save

a quantity of data and prolong the video/audio recording time, can be implemented.

In order to achieve the above object, an audio information transforming device set forth in claim 12 provides an audio information transforming device for a video/audio format in which a scene reproduced on a screen is constructed to contain objects and each object has video information, position information, and audio information, which comprises a means for deciding a virtual listening point at a position different from a basic listening point that is set as a position at which a listener listens to an audio; a means for comparing a positional relationship between the basic listening point and the object with a positional relationship between the virtual listening point and the object; and a means for changing an allocation ratio of an audio to a plurality of audio outputting means based on a result of the comparing means.

According to such device, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based

on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced by employing the above  
5 audio information transforming device.

Also, an audio information transforming device set forth in claim 13 provides an audio information transforming device for a video/audio format in which each  
10 scene produced on a screen has video information, audio information, and a virtual sound source, which comprises a means for deciding a virtual listening point at a position different from a basic listening point that is set as a position at which a listener listens to an audio; a means  
15 for comparing a positional relationship between the basic listening point and the virtual sound source with a positional relationship between the virtual listening point and the virtual sound source; and a means for changing an allocation ratio of an audio to a plurality of audio  
20 outputting means based on a result of the comparing means.

According to such device, the positional relationship between the basic listening point and the virtual sound source is calculated and also the positional relationship between the virtual listening point and the  
25 virtual sound source is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means

(e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced by employing the above audio information transforming device.

Also, an audio information transforming device set forth in claim 14 provides an audio information transforming device for a video/audio format in which a screen contains a plurality of objects and each object has video information, position information, and 1-channel audio information, which comprises a means for calculating a positional relationship between a basic listening point, which is set as a position at which a listener listens to an audio, and the object; and a means for allocating the 1-channel audio information to a plurality of audio outputting means based on the positional relationship.

According to such device, the 1-channel audio information that the object has can be allocated to the multiple channels of the audio based on the positional relationship between the basic listening point and the object, and can be output to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system). Therefore, if the above audio information transforming device is employed, a quantity of data can be

saved and the video/audio recording time can be prolonged.

Also, an audio information transforming device set forth in claim 15 provides an audio information transforming device for a video/audio format in which a screen contains a plurality of objects and each object has video information, position information, and 1-channel audio information, which comprises a means for deciding a virtual listening point at a position different from a basic listening point that is set as a position at which a listener listens to an audio; a means for comparing a positional relationship between the basic listening point and the object with a positional relationship between the virtual listening point and the object; and a means for allocating the 1-channel audio information to a plurality of audio outputting means based on a result of the comparing means.

According to such device, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the 1-channel audio information can be allocated to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated

positional relationships. Therefore, if the above audio information transforming device is employed, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (virtual listening point), can be produced, and also a quantity of data can be saved and the video/audio recording time can be prolonged.

In addition, in the audio information transforming device according to any one of claim 12, claim 13, or claim 15 as the audio information transforming device set forth in claims 16 to 18, the virtual listening point or the virtual sound source has direction information.

According to such device, the direction information is provided to the virtual listening point or the virtual sound source. Therefore, if the above audio information transforming device is employed, the listener can feel that such listener just enters into the video (virtual listening point), and the direction to which the listener looks can be changed, and the audio environment with the stronger appeal/reality can be produced.

#### **Brief Description of the Drawings**

FIG.1 is a view explaining a first embodiment according to the present invention;

FIG.2 is a view explaining the first embodiment



according to the present invention, and a view showing positional images among respective loudspeakers, a basic listening point, and a virtual listening point in an X-Z coordinate plane;

5           FIG.3 is a view explaining the first embodiment according to the present invention, and a view showing positional images among a virtual sound source, a basic listening point, and a virtual listening point in an X-Z coordinate plane;

10           FIG.4 is a view explaining the first embodiment according to the present invention, and a view showing a positional relationship between the virtual listening point and the virtual sound source when such virtual listening point is regarded as a center;

15           FIG.5 is a view explaining a second embodiment according to the present invention;

            FIG.6 is a view explaining a third embodiment according to the present invention;

20           FIG.7 is a view explaining the third embodiment according to the present invention, and a view showing a positional relationship between the basic listening point and an object when such basic listening point is regarded as the center;

25           FIG.8 is a view explaining a fourth embodiment according to the present invention, and a view showing a positional relationship between the virtual listening point

and the object when such virtual listening point is regarded as the center; and

FIG.9 is a view explaining a fifth embodiment according to the present invention;

5

In the drawings, the reference numerals 101, 501, and 601, each refers to a screen; 102 to a basic listening point; each of 103, 104, and 602 to an object; 105 to a virtual listening point; each of 502 and 901 to a  
10 virtual sound source; and each of 110 to 115 to a loudspeaker.

#### Detailed Description of the Preferred Embodiments

Embodiments according to the present invention will  
15 be explained in detail with reference to the drawings hereinafter. Details of an audio information transforming method will be explained in detail in respective embodiments hereinafter. In this case, this audio information transforming method can be implemented as an  
20 audio information transforming program as procedures that are executed by a computer, and can be constructed as an audio information transforming device by storing this audio information transforming program in a storing medium such as a memory, or the like.

25

(First Embodiment)

A first embodiment according to the present invention will be explained with reference to FIG.1 to FIG.4 hereunder.

5           The present embodiment is concerned with a video/audio format such as MPEG 4, or the like, in which the screen contains a plurality of objects and each object has video information, position information, and audio information. Assume that a screen 101 shown in FIG.1 has  
10   the audio information that is supposed to be listened by the listener at a basic listening point 102.

Also, assume that the scene is being played on the screen 101 and two objects 103, 104 are present in the scene. The objects 103, 104 have video stream (video  
15   information), audio stream (audio information), and position information respectively.

The virtual listening point 105 is decided on the coordinate in such screen 101, and then the audio environment that permits the listener to feel that such  
20   listener is listening to the audio just at a virtual listening point 105 is produced.

As an example of the multi-channel audio system, an example of 5.1-channel will be explained herein.

The object 103 has the audio stream in 5.1-channels.  
25   Assume that the audio information in respective channels are Ob1L, Ob1R, Ob1C, Ob1Ls, Ob1Rs, Ob1Lfe, as shown in

FIG.2, for example.

Here, an R-channel of stream data of the object 103, for example, is observed. Assume that the object 103 itself has a sound source and that, as shown in FIG.3, a virtual sound source of the R component (audio information 0b1R) of the object 103 is S(k,k).

Also, as shown in FIG.1, assume that a distance from the basic listening point 102 to the object 103 is set to L1 and a distance from the virtual listening point 105 to the object 103 is set to L2. Then, a loudness Vvr1 of the sound, which is listened at the virtual listening point 105, with respect to a loudness Vbr1 of the sound in the R channel, which is listened at the basic listening point 102, is calculated by Equation (1) based on a ratio of the distance L1 between the basic listening point 102 and the object 103 to the distance L2 between the virtual listening point 105 and the object 103. In this case, assume that the loudness of the sound is in inverse proportion to square of the distance.

[Formula 1]

$$V_{vr1} = \frac{l_1^2}{l_2^2} V_{br1} \quad \dots (1)$$

The loudness Vvr1 of the sound that is listened at the virtual listening point 105 is regarded as the loudness of the sound emitted from the virtual sound source S, and then the sound is allocated to respective loudspeakers 110,

111, 113, and 114. In other words, as shown in FIG.4, a circle on a circumference of which the virtual sound source S is positioned is supposed around the virtual listening point 105 as the center. Then, a perpendicular is drawn from the virtual sound source S to the X-axis and the Z-axis respectively, then intersection points between the perpendiculars and respective axes and also intersection points between the circle and respective axes are detected, and then ratios of their distances are calculated.. Then, if calculated results are represented as T1:T2 and U1:U2 respectively, an allocation ratio of the sound that is emitted from the virtual sound source S is given as follows.

[Formula 2]

$$L:R=T2^2:T1^2$$

[Formula 3]

$$Front:Rear=U2^2:U1^2$$

That is, a relationship is given as follows. The loudness Vvr1 of the sound in the R channel, which is listened at the virtual listening point 105, is allocated to respective loudspeakers 110, 111, 113, and 114 based on this ratio.

[Formula 4]

$$L:R:Ls:Rs=T2^2:T1^2:\frac{U1^1}{U2^2}T2^2:\frac{U1^2}{U2^2}T1^2$$

That is to say, the R component obtained when the sound of the object 103 in the R channel with respect to

the basic listening point 102 is listened at the virtual listening point 105 can be derived by Equation (2).

[Formula 5]

$$V_{v\_r\_r1} = \frac{T1^2}{T1^2 + T2^2 + \frac{U1^2}{U2^2} T1^2 + \frac{U1^2}{U2^2} T2^2} V_{vr1} = \frac{T1^2}{(1 + \frac{U1^2}{U2^2})(T1^2 + T2^2)} \frac{l1^2}{l2^2} V_{br1}$$

...(2)

In this case, above Equation (2) indicates the R component obtained when the R-channel component of the object 103 (R-channel component at the basic listening point 102) is allocated to the audio component that is listened at the virtual listening point 105.

Also, the loudness  $V_{vr1}$  of the sound that is listened at the virtual listening point 105 is allocated to L component, Rs component, and Ls component in the same way as the above calculation. The data of the object 103 except the R channel are allocated similarly. Then, the audio information when the sound of the object 103 is listened at the virtual listening point 105 are completed by adding these data. The R component, which is taken as an example, is given by an equation (3).

[Formula 6]

$$V_{v\_r1} = V_{v\_r\_r1} + V_{v\_r\_l1} + V_{v\_r\_rs1} + V_{v\_r\_ls1} \quad \dots (3)$$

The audio information is completed every scene by executing these calculations every object and then adding calculated results. The R component, which is taken as an

example, is given by a following equation.

[Formula 7]

$$Vv_r = Vv_{r1} + Vv_{r2} + Vv_{r3} + \dots \quad \dots (4)$$

Also, the above calculation is applied to L component, Rs  
5 component, and Ls component. Thus, the audio information,  
which are to be listened at the virtual listening point  
105, are completed by outputting the audio information onto  
respective channels.

In this manner, according to the present  
10 embodiment, the virtual listening point 105 is decided at  
the position that is different from the basic listening  
point 102, then the positional relationship between the  
virtual listening point 105 and the object and the  
positional relationship between the basic listening point  
15 102 and the object are calculated every object, and then an  
allocation ratio of the audio to the loudspeakers 110, 111,  
113 and 114 is changed based on these positional  
relationships. Therefore, the audio environment with  
strong appeal/reality, which enables the listener to feel  
20 that such listener is positioned just at the virtual  
listening point 105 in the video, can be produced.

In this case, in the present embodiment, an example  
of the 5.1-channel system is explained as the multi-channel  
audio system. But other audio systems each having the  
25 different channel number may be employed.

(Second Embodiment)

Next, a second embodiment according to the present invention will be explained with reference to FIG.5 hereunder.

5 Unlike the case (MPEG 4, or the like) in which, as shown in the above first embodiment, the screen has the video information and the audio information in unit of the object, the case in which the screen has the video information and the audio information in unit of the scene, 10 which is reproduced on the screen like the existing DVD, will be explained herein.

In other words, assume that the scene is reproduced in a screen 501 shown in FIG.5 and this scene has the video information and the audio information that is supposed to 15 be listened at the basic listening point 102. An example in which the audio information that are to be output from respective channels are prepared when the virtual listening point 105 is decided will be explained.

First, a virtual sound source 502 is decided. The 20 calculation method in the first embodiment can be employed by replacing the object in the first embodiment with this virtual sound source 502. Assume that the virtual sound source 502 is decided instead of the object in FIG.1 and a distance from the basic listening point 102 is set to L1 and a distance from the virtual listening point 105 is set 25 to L2. As shown in FIG.3, the virtual sound source S (k,k)



is decided in the direction to the loudspeaker. Then, as shown in FIG.4, the circle on the circumference of which the virtual sound source S is positioned is supposed around the virtual listening point 105 as the center. If a  
 5 portion ( $V_{v\_r\_r}$ ) that is listened at the virtual listening point 105 is calculated by allocating the channel R components  $V_{br}$ ,  $V_{br}$ , which are listened at the basic listening point 102, such portion is given by an equation (5).

10 [Formula 8]

$$V_{v\_r\_r} = \frac{T1^2}{T1^2 + T2^2 + \frac{U1^2}{U2^2} T1^2 + \frac{U1^2}{U2^2} T2^2} V_{vr} = \frac{T1^2}{(1 + \frac{U1^2}{U2^2})(T1^2 + T2^2)} \frac{l1^2}{l2^2} V_{br}$$

... (5)

The R component of the sound, which is listened at  
 15 the virtual listening point 105, is completed by adding all the allocated R components.

[Formula 9]

$$V_{v\_r} = V_{v\_r\_r} + V_{v\_r\_l} + V_{v\_r\_rs} + V_{v\_r\_ls} \quad \dots (6)$$

In this case, L component, Rs component, and Rl  
 20 component of the sound, which is listened at the basic listening point 102, except the R component are also allocated to the sounds that are to be listened at the virtual listening point 105, by the similar calculation.

When the data are used as the audio information and  
 25 then the audio in the data are output from respective loudspeakers 110, 111, 113, and 114 as the audio

information, the listener can enjoy the audio environment in which the virtual listening point 105 is set as the listening point. Also, the listener can enjoy many audio environments by changing the coordinate system of the  
5 virtual listening point 105.

In this fashion, according to the present embodiment, the virtual listening point 105 is decided at the position that is different from the basic listening point 102, then the positional relationship between the  
10 virtual listening point 105 and the virtual sound source 502 and the positional relationship between the basic listening point 102 and the virtual sound source 502 are calculated every scene, and then an allocation ratio of the audio to the loudspeakers 110, 111, 113 and 114 is changed  
15 based on these positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener is positioned just at the virtual listening point 105 in the video, can be produced.

20

#### (Third Embodiment)

A third embodiment according to the present invention will be explained with reference to FIG.6 and FIG.7 hereunder.

25 In the present embodiment, assume that the scene that is reproduced on the screen consists of the background

and objects and each object has the video stream and the audio stream in which 1-channel audio information are contained. As shown in FIG.6, the basic listening point 102 is set. Several objects (respective audio streams have the one-channel audio information) are present on the scene. For example, the audio information of an object 602 are produced by the basic listening point 102 and the coordinate information of the object 602 as the multi-channel audio information.

As shown in FIG.7, a circle is drawn such that the circle has the basic listening point 102 as the center and also the object 602 is positioned on its circumference. Then, a perpendicular is drawn from the object 602 to the X-axis and the Z-axis respectively, and then intersection points between the perpendiculars and the X-axis and the Z-axis and also intersection points between the circle and the X-axis and the Z-axis are detected respectively. Then, if ratios of distances from the intersection points between the circle and the concerned axis to the intersection point between the perpendicular and the concerned axis are represented as the X-axis= $T1:T2$  and the Z-axis= $U1:U2$  respectively, an allocation ratio of the sound that is emitted from the object 602 is given as follows.

[Formula 10]

$$L:R=T2^2:T1^2$$

[Formula 11]

$$Front:Rear = U2^2 : U1^2$$

That is, the allocation ratio is given as follows.

[Formula 12]

$$L:R:Rs:Rs = T2^2:T1^2:\frac{U1^1}{U2^2}T2^2:\frac{U1^2}{U2^2}T1^2$$

5       The audio information in the audio stream, that the  
object 602 has, are allocated to respective channels based  
on this ratio. When a plurality of objects are present,  
this process is executed every object, and then allocated  
audio information in respective channels are added and  
10   output from respective loudspeakers 110, 111, 113 and 114.  
Therefore, the listener can enjoy the multi-channel audio  
even in the scene that is composed of the object that has  
only the 1-channel audio information. Also, since there is  
no necessity that the multi-channel data should be loaded  
15   previously, a quantity of data can be reduced.

In this manner, according to the present  
embodiment, in the video/audio format that has video  
information, position information of this video data, and  
1-channel audio information every object, the allocation  
20   ratio of the audio to the loudspeakers 110, 111, 113, and  
115 is changed based on the positional relationship between  
the basic listening point 102 and the object every object.  
Therefore, the reality can be enhanced by producing the  
multi-channel audio information even from the 1-channel  
25   audio information, and the audio information can be reduced  
to one channel only.

(Fourth Embodiment)

A fourth embodiment according to the present invention will be explained with reference to FIG.6 and  
5 FIG.8 hereunder.

The virtual listening point 105 is added to the configuration in which the object 602 has the 1-channel audio information, as described in the above third embodiment. . . . The audio environment that permits the  
10 listener to feel that such listener is listening to the sound just at the virtual listening point 105 can be produced by adding the virtual listening point 105.

In FIG.6, assume that a distance between the basic listening point 102 and the object 602 is  $L1$  and a distance  
15 between the virtual listening point 105 and the object 602 is  $L2$ . In FIG.8, a circle is drawn such that the circle has the virtual listening point 105 as the center and also the object 602 is positioned at a point on its circumference. Then, a perpendicular is drawn from the  
20 object 602 to the X-axis and the Z-axis respectively, and then intersection points between the perpendiculars and the X-axis and the Z-axis and also intersection points between the circle and the X-axis and the Z-axis are calculated respectively. Then, if ratios of distances from the  
25 intersection points between the circle and the concerned axis to the intersection point between the perpendicular

and the concerned axis are represented as the X-axis= $T1:T2$  and the Z-axis= $U1:U2$  respectively, an allocation ratio of the sound that is emitted from the object 602 is given as follows.

5 [Formula 13]

$$L:R=T2^2:T1^2$$

[Formula 14]

$$Front:Rear=U2^2:U1^2$$

That is, the allocation ratio is given as follows.

10 [Formula 15]

$$L:R:Ls:Rs=T2^2:T1^2:\frac{U1^1}{U2^2}T2^2:\frac{U1^2}{U2^2}T1^2$$

If the audio information of the object 602 is assumed as  $Vobj1$ , the R component, which is taken as an example, is given by an equation (7).

15 [Formula 16]

$$V_{v\_r1} = \frac{T1^1}{1 + (\frac{U1^2}{U2^2})(T1^2 + T2^2)} \frac{l1^2}{l2^2} Vobj1 \quad \dots (7)$$

The R component of the audio information, which is to be listened at the virtual listening point, can be produced by deciding the allocation of the R components of the audio information emitted from all objects based on the similar calculation and then adding these components. Also, the audio information, which is to be listened at the virtual listening point, can be produced by applying the above calculation to remaining L component, Rs component,

20

25

and Ls component.

In this manner, according to the present embodiment, in the video/audio format that has video information, position information of this video data, and  
5 1-channel audio information every object, the distance L1 between the basic listening point 102 and the object 602 and the distance L2 between the virtual listening point 105 and the object 602 are calculated, and then the allocation ratio of the audio to the loudspeakers 110, 111, 113, and  
10 114 is changed based on the result. Therefore, the reality can be enhanced by producing the multi-channel audio information even from the 1-channel audio information, and the audio information can be reduced to one channel only.

15 (Fifth Embodiment)

Next, a fifth embodiment according to the present invention will be explained with reference to FIG.9 hereunder.

The listener's choice can be further increased by providing  
20 the direction information to the virtual listening point 105 in the first embodiment, the second embodiment, and the fourth embodiment. More particularly, the audio environment that permits the listener to feel that such listener just stands on the virtual listening point 105 and  
25 looks to the direction indicated by the direction information of the virtual listening point 105 is produced

by changing the allocation of the audio information to respective channels. Information that the virtual listening point 105 has are the coordinate and the direction vector.

5           When the positional relationship between the virtual sound source  $S(k,k)$  and the virtual listening point 105 is considered in the first embodiment and the second embodiment and also the positional relationship between the object 602 and the virtual listening point 105 is  
10 considered in the fourth embodiment, the direction information of the virtual listening point 105 is also considered.

In the first embodiment, the second embodiment, and the fourth embodiment, the allocation of the audio  
15 information is executed on the premise that the listener looks to the front in the Z-axis direction. Therefore, when the direction of the virtual sound source (the first embodiment and the second embodiment) or the object 602 (the fourth embodiment) is considered, an angle between the  
20 Z-axis and the direction vector is considered. In this example, the rotation on the X-Z plane is considered.

As shown in FIG.9, the X'-axis and the Z'-axis are set by rotating the X-axis and the Z-axis by an angle  $\theta$  between the direction vector and the Z-axis. Then, a  
25 perpendicular is drawn from the virtual sound source  $S(k,k)$  the object 602 to the X'-axis and the Z'-axis



respectively, and then intersection points between the perpendiculars and the X'-axis and the Z'-axis and also intersection points between the circle and the X'-axis and the Z'-axis are calculated respectively. Then, ratios of distances from the intersection points between the circle and the concerned axis to the intersection point between the perpendicular and the concerned axis are represented as the X'-axis=T1:T2 and the Z'-axis=U1:U2 respectively. Subsequently, according to the calculation method explained in the first embodiment, the second embodiment, and the fourth embodiment, the audio information can be supplied to respective channels in the audio environment in which the listener looks to the direction indicated by the direction vector from the basic listening point 102.

In this manner, according to the present embodiment, the direction vector information as well as the position information is provided to the virtual listening point 105, then the distance/direction to the virtual listening point are calculated based on the scene coordinate information that respective objects have, and then the allocation ratio of the audio to the loudspeakers 110, 111, 113, and 114 is changed based on the result. Therefore, the reality can be enhanced by producing the multi-channel audio information even from the 1-channel audio information. In addition, since the direction vector information as well as the position information is provided

to the virtual listening point 105, the reality can be enhanced further rather than the first embodiment, the second embodiment, and the fourth embodiment, described above.

5           If a decoder for decoding the video/audio format is constructed by using the audio information transforming method, the audio information transforming program, the audio information transforming device in respective  
... .. embodiments described above, the video/audio player (DVD ... ..  
10 player, LD player, MPEG player, game, system in the movie theater, etc.), which can achieve advantages of respective embodiments, can be implemented.

### **Advantages of the Invention**

15           As described in detail above, according to the audio information transforming method set forth in Claim 1, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is  
20 calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker  
25 system) is changed based on the compared result of

respective calculated positional relationships. Therefore,  
the audio environment with the strong appeal/reality, which  
enables the listener to feel that such listener just enters  
into the video (the virtual listening point), can be  
5 produced.

Also, according to the audio information  
transforming method set forth in Claim 2, the positional  
relationship between the basic listening point and the  
virtual sound source is calculated and also the positional  
10 relationship between the virtual listening point and the  
virtual sound source is calculated, and then the allocation  
ratio of the audio to a plurality of audio outputting means  
(e.g., the 5.1-channel loudspeaker system) is changed based  
on the compared result of respective calculated positional  
15 relationships. Therefore, the audio environment with the  
strong appeal/reality, which enables the listener to feel  
that such listener just enters into the video (the virtual  
listening point), can be produced.

Also, according to the audio information  
20 transforming method set forth in Claim 3, the 1-channel  
audio information that the object has can be allocated to  
the multiple channels of the audio based on the positional  
relationship between the basic listening point and the  
object, and can be output to a plurality of audio  
25 outputting means (e.g., the 5.1-channel loudspeaker  
system). Therefore, a quantity of data can be saved and

the video/audio recording time can be prolonged.

Also, according to the audio information transforming method set forth in Claim 4, the virtual listening point is decided at the position that is  
5 different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and  
10 then the 1-channel audio information can be allocated to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which  
15 enables the listener to feel that such listener just enters into the video (virtual listening point), can be produced, and also a quantity of data can be saved and the video/audio recording time can be prolonged.

Also, according to the audio information  
20 transforming method set forth in claims 5 to 7, the direction information is provided to the virtual listening point or the virtual sound source. Therefore, the listener can feel that such listener just enters into the video (virtual listening point), and the direction to which the  
25 listener looks can be changed, and the audio environment with the stronger appeal/reality can be produced.

According to the audio information transforming program set forth in claim 8, in the reproduced audio having the video/audio format such as MPEG 4, for example, the positional relationship between the basic listening point and the virtual sound source is calculated and also the positional relationship between the virtual listening point and the virtual sound source is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio environment with the strong appeal/reality, which permits the listener to feel that such listener just enters into the video (the virtual listening point), can be implemented.

Also, according to the audio information transforming program set forth in claim 9, in the video/audio format having the audio information and the video information in unit of the scene that is reproduced on the screen, the positional relationship between the basic listening point and the virtual sound source is calculated and also the positional relationship between the

virtual listening point and the virtual sound source is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio environment with the strong appeal/reality, which permits the listener to feel that such listener just enters into the video (the virtual listening point), can be implemented.

Also, according to the audio information transforming program set forth in claim 10, in the reproduced audio having the video/audio format such as MPEG 4, for example, the 1-channel audio information that the object has can be allocated to the multiple channels of the audio based on the positional relationship between the basic listening point and the object, and can be output to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system). Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can save a quantity of data and

prolong the video/audio recording time, can be implemented.

Also, according to the audio information transforming program set forth in claim 11, in the reproduced audio having the video/audio format such as MPEG  
5 4, for example, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional  
10 relationship between the object and the basic listening point is calculated, and then the 1-channel audio information is allocated to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional  
15 relationships. Therefore, if the recording medium (the memory such as ROM, or the like) in which this program is recorded is employed, the video/audio player (DVD player, LD player, MPEG player, game, system in the movie theater, etc.), which can produce the audio environment with the  
20 strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), and also can save a quantity of data and prolong the video/audio recording time, can be implemented.

According to the audio information transforming  
25 device set forth in claim 12, the virtual listening point is decided at the position that is different from the basic

listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced by employing the above audio information transforming device.

Also, according to the audio information transforming device set forth in claim 13, the positional relationship between the basic listening point and the virtual sound source is calculated and also the positional relationship between the virtual listening point and the virtual sound source is calculated, and then the allocation ratio of the audio to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) is changed based on the compared result of respective calculated positional relationships. Therefore, the audio environment with the strong appeal/reality, which enables the listener to feel that such listener just enters into the video (the virtual listening point), can be produced by employing the above



audio information transforming device.

Also, according to the audio information transforming device set forth in claim 14, the 1-channel audio information that the object has can be allocated to the multiple channels of the audio based on the positional relationship between the basic listening point and the object, and can be output to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system). Therefore, if the above audio information transforming device is employed, a quantity of data can be saved and the video/audio recording time can be prolonged.

Also, according to the audio information transforming device set forth in claim 15, the virtual listening point is decided at the position that is different from the basic listening point, then the positional relationship to the virtual listening point is calculated based on the position information that the object has and also the positional relationship between the object and the basic listening point is calculated, and then the 1-channel audio information can be allocated to a plurality of audio outputting means (e.g., the 5.1-channel loudspeaker system) based on the compared result of respective calculated positional relationships. Therefore, if the above audio information transforming device is employed, the audio environment with the strong appeal/reality, which enables the listener to feel that

such listener just enters into the video (virtual listening point), can be produced, and also a quantity of data can be saved and the video/audio recording time can be prolonged.

In addition, according to the audio information transforming device set forth in claims 16 to 18, the direction information is provided to the virtual listening point or the virtual sound source. Therefore, if the above audio information transforming device is employed, the listener can feel that such listener just enters into the video (virtual listening point), and the direction to which the listener looks can be changed, and the audio environment with the stronger appeal/reality can be produced.